

# Slow Down & Sleep for Profit in Online Deadline Scheduling<sup>\*</sup>

Andreas Cord-Landwehr, Peter Kling, and Frederik Mallmann-Trenn

Heinz Nixdorf Institute and Computer Science Department, University of Paderborn  
{andreas.cord-landwehr@, peter.kling@, xarph@mail.}uni-paderborn.de

**Abstract.** We present and study a new model for energy-aware and profit-oriented scheduling on a single processor. The processor features dynamic speed scaling as well as suspension to a sleep mode. Jobs arrive over time, are preemptable, and have different sizes, values, and deadlines. On the arrival of a new job, the scheduler may either accept or reject the job. Accepted jobs need a certain energy investment to be finished in time, while rejected jobs cause costs equal to their values. Here, power consumption at speed  $s$  is given by  $P(s) = s^\alpha + \beta$  and the energy investment is power integrated over time. Additionally, the scheduler may decide to suspend the processor to a sleep mode in which no energy is consumed, though awaking entails fixed transition costs  $\gamma$ . The objective is to minimize the total value of rejected jobs plus the total energy.

Our model combines aspects from advanced energy conservation techniques (namely speed scaling and sleep states) and profit-oriented scheduling models. We show that *rejection-oblivious* schedulers (whose rejection decisions are not based on former decisions) have – in contrast to the model without sleep states – an unbounded competitive ratio w.r.t. the processor parameters  $\alpha$  and  $\beta$ . It turns out that the worst-case performance of such schedulers depends linearly on the jobs’ value densities (the ratio between a job’s value and its work). We give an algorithm whose competitiveness nearly matches this lower bound. If the maximum value density is not too large, the competitiveness becomes  $\alpha^\alpha + 2e\alpha$ . Also, we show that it suffices to restrict the value density of low-value jobs only. Using a technique from [12] we transfer our results to processors with a fixed maximum speed.

## 1 Introduction

Over the last decade, energy usage of data centers and computers in general has become a major concern. There are various reasons for this development: the ubiquity of technical systems, the rise of mobile computing, as well as a growing ecological awareness. Also from an economical viewpoint, energy usage can no longer be ignored. Energy costs for both the actual computation and the cooling have become *the* decisive cost factor in today’s data centers (see, e.g., Barroso and Hölzle [10]). In combination with improvements on the technical level, algorithmic research has great potential to reduce energy consumption. Albers [2] gives a good insight on the role of algorithms to fully exploit the energy-saving mechanisms of modern systems. Two of the most prominent techniques for power saving are *dynamic speed scaling* and *power-down*. The former allows a system to save energy by adapting the processor’s speed to the current system load, while the latter can be used to transition into a sleep mode to conserve energy. There is an extensive body of literature on both techniques (see below). From an algorithmic viewpoint, the most challenging aspect in the design of scheduling strategies is to handle the lack of knowledge

---

<sup>\*</sup> This work was partially supported by the German Research Foundation (DFG) within the Collaborative Research Center “On-The-Fly Computing” (SFB 901) and by the Graduate School on Applied Network Science (GSANS).

about the future: should we use a high speed to free resources in anticipation of new jobs or enter sleep mode in the hope that no new jobs arrive in the near future?

Given that profitability is a driving force for most modern systems and that energy consumption has gained such a high significance, it seems natural to take this relation explicitly into account. Pruhs and Stein [16] consider a scheduling model that does so by introducing job values. Their scheduler controls energy usage via speed scaling and is allowed to reject jobs if their values seem too low compared to their foreseeable energy requirements. The objective is to maximize the profit, which is modeled as the total value of finished jobs minus the invested energy. Our work is based on a result by Chan, Lam, and Li [12]. We enhance their model by combining speed scaling and power-down mechanisms for energy management, which not only introduces non-trivial difficulties to overcome in the analysis, but proves to be inherently more complex compared to the original model insofar that classical algorithms can become arbitrarily bad.

**History & Related Work.** There is much literature concerning energy-aware scheduling strategies both in practical and theoretical contexts. A recent survey by Albers [1] gives a good and compact overview on the state of the art in the dynamic speed scaling setting, also in combination with power-down mechanisms. In the following, we focus on theoretical results concerning scheduling on a single processor for jobs with deadlines. Theoretical work in this area has been initiated by Yao, Demers, and Shenker [17]. They considered scheduling of jobs having different sizes and deadlines on a single variable-speed processor. When running at speed  $s$ , its power consumption is  $P(s) = s^\alpha$  for some constant  $\alpha \geq 2$ . Yao, Demers, and Shenker derived a polynomial time optimal offline algorithm as well as two online algorithms known as *optimal available* (OA) and *average rate* (AVR). Up to now, OA remains one of the most important algorithms in this area, as it is used as a basic building block by many strategies (including the strategy we present in this paper). Using an elegant amortized potential function argument, Bansal, Kimbrel, and Pruhs [5] were able to show that OA’s competitive factor is exactly  $\alpha^\alpha$ . Moreover, the authors stated a new algorithm, named BKP, which achieves a competitive ratio of essentially  $2e^{\alpha+1}$ . This improves upon OA for large  $\alpha$ . The best known lower bound for deterministic algorithms is  $e^{\alpha-1}/\alpha$  due to Bansal et al. [6]. They also presented an algorithm (qOA) that is particularly well-suited for low powers of  $\alpha$ . An interesting and realistic model extension is the restriction of the maximum processor speed. In such a setting, a scheduler may not always be able to finish all jobs by their deadlines. Chan et al. [13] were the first to consider the combination of classical speed scaling with such a maximum speed. They gave an algorithm that is  $\alpha^\alpha + \alpha^2 4^\alpha$ -competitive on energy and 14-competitive on throughput. Bansal et al. [7] improved this to a 4-competitive algorithm concerning the throughput while maintaining a constant competitive ratio with respect to the energy. Note that no algorithm – even if ignoring the energy consumption – can be better than 4-competitive for throughput (see [11]).

Power-down mechanisms were studied by Baptiste [8]. He considered a fixed-speed processor needing a certain amount of energy to stay awake, but which may switch into a sleep state to save energy. Returning from sleep needs energy  $\gamma$ . For jobs of unit size, he gave a polynomial time optimal offline algorithm, which was later extended to jobs of arbitrary size [9]. The first work to combine both dynamic speed scaling and sleep states in the classical YAO-model is due to Irani, Shukla, and Gupta [15]. They achieved a 2-approximation for arbitrary convex power functions. For the online setting and power function  $P(s) = s^\alpha + \beta$  a competitive factor of  $4^{\alpha-1}\alpha^\alpha + 2^{\alpha-1} + 2$  was reached. Han et al. [14] improved upon this in two respects: they lowered the competitive factor to  $\alpha^\alpha + 2$  and transferred the result to scenarios limiting the maximum speed. Only recently, Albers and Antoniadis [3] proved that the optimization problem is NP-hard and gave lower bounds for several algorithm classes. Moreover, they improved the approximation factor for general convex power functions to  $4/3$ . The papers most closely related to ours are

due to Pruhs and Stein [16] and Chan, Lam, and Li [12]. Both considered the dynamic speed scaling model of Yao, Demers, and Shenker. However, they extended the idea of energy-minimal schedules to a profit-oriented objective. In the simplest case, jobs have values (or priorities) and the scheduler is no longer required to finish all jobs. Instead, it can decide to reject jobs whose values do not justify the foreseeable energy investment necessary to complete them. The objective is to maximize profit [16] or, similarly, minimize the loss [12]. As argued by the authors, the latter model has the benefit of being a direct generalization of the classical model of Yao, Demers, and Shenker [17]. For maximizing the profit, Pruhs and Stein [16] showed that, in order to achieve a bounded competitive factor, resource augmentation is necessary and gave a scalable online algorithm. For minimizing the loss, Chan, Lam, and Li [12] gave a  $\alpha^\alpha + 2e\alpha$ -competitive algorithm and transferred the result to the case of a bounded maximum speed.

**Our Contribution.** We present the first model that not only takes into account two of the most prominent energy conservation techniques (namely, speed scaling and power-down) but couples the energy minimization objective with the idea of profitability. It combines aspects from both [15] and [12]. From [15] we inherit one of the most realistic processor models considered in this area: A single variable-speed processor with power function  $P(s) = s^\alpha + \beta$  and a sleep state. Thus, even at speed zero the system is charged a certain amount  $\beta$  of energy, but it can suspend to sleep such that no energy is consumed. Waking up causes transition cost of  $\gamma$ . The job model stems from [12]: Jobs arrive in an online fashion, are preemptable, and have a deadline, size, and value. The scheduler can reject jobs (e.g., if their values do not justify the presumed energy investment). Its objective is to minimize the total energy investment plus the total value of rejected jobs.

A major insight of ours is that the maximum value density  $\delta_{\max}$  (i.e., the ratio between a job's value and its work) is a parameter that is inherently connected to the necessary and sufficient competitive ratio achievable for our online scheduling problem. We present an online algorithm that combines ideas from [12] and [14] and analyze its competitive ratio with respect to  $\delta_{\max}$ . This yields an upper bound of  $\alpha^\alpha + 2e\alpha + \delta_{\max} \frac{s_{\text{cr}}}{P(s_{\text{cr}})}$ .<sup>1</sup> If the value density of low-valued jobs is not too large or job values are at least  $\gamma$ , the competitive ratio becomes  $\alpha^\alpha + 2e\alpha$ . Moreover, we show that one cannot do much better: any *rejection-oblivious* strategy has a competitive ratio of at least  $\delta_{\max} \frac{s_{\text{cr}}}{P(s_{\text{cr}})}$ . Here, rejection-oblivious means that rejection decisions are based on the *current* system state and job properties only. This lower bound is in stark contrast to the setting without sleep states, where a rejection-oblivious  $O(1)$ -competitive algorithm exists [12]. Using the definition of a job's penalty ratio (due to Chan, Lam, and Li [12]), we extend our results to processors with a bounded maximum speed.

## 2 Model & Preliminaries

We are given a speed-scalable processor that can be set to any speed  $s \in [0, \infty)$ . When running at speed  $s$  its power consumption is  $P_{\alpha,\beta}(s) = s^\alpha + \beta$  with  $\alpha \geq 2$  and  $\beta \geq 0$ . If  $s(t)$  denotes the processor speed at time  $t$ , the total power consumption is  $\int_0^\infty P_{\alpha,\beta}(s(t)) dt$ . We can suspend the processor into a sleep state to save energy. In this state, it cannot process any jobs and has a power consumption of zero. Though entering the sleep state is for free, waking up needs a fixed *transition energy*  $\gamma \geq 0$ . Over time,  $n$  jobs  $J = \{1, 2, \dots, n\}$  are released. Each job  $j$  appears at its release time  $r_j$  and has a deadline  $d_j$ , a (non-negative) value  $v_j$ , and requires a certain amount  $w_j$  of work. The processor can process at most one job at a time. Preemption is allowed, i.e., jobs may be paused at any time and continued later on. If  $I$  denotes the period of

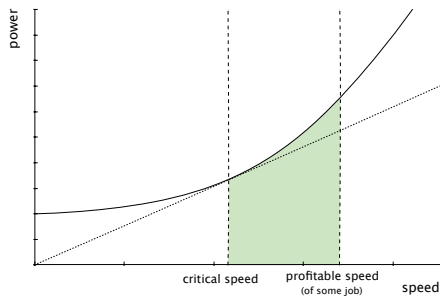
<sup>1</sup> The expression  $\frac{s_{\text{cr}}}{P(s_{\text{cr}})}$  depends only on  $\alpha$  and  $\beta$ , see Section 2.

time (not necessarily an interval) when  $j$  is scheduled, the amount of work processed is  $\int_I s(t) dt$ . A job is finished if  $\int_I s(t) dt \geq w_j$ . Jobs not finished until their deadline cause a cost equal to their value. We call such jobs *rejected*. A schedule  $S$  specifies for any time  $t$  the processor's state (asleep or awake), the currently processed job (if the processor is awake), and sets the speed  $s(t)$ . W.l.o.g. we assume  $s(t) = 0$  when no job is being processed. Initially, the processor is assumed to be asleep. Whenever it is neither sleeping nor working we say it is *idle*. A schedule's cost is the invested energy (for awaking from sleep, idling, and working on jobs) plus the loss due to rejected jobs. Let  $m$  denote the number of sleep intervals,  $l$  the total length of idle intervals, and  $\mathcal{I}_{\text{work}}$  the collection of all working intervals (i.e., times when  $s(t) > 0$ ). Then, the schedule's *sleeping energy* is  $E_{\text{sleep}}^S := (m-1)\gamma$ , its *idling energy* is  $E_{\text{idle}}^S := l\beta$ , and its *working energy* is  $E_{\text{work}}^S := \int_{\mathcal{I}_{\text{work}}} P_{\alpha,\beta}(s(t)) dt$ . We use  $V_{\text{rej}}^S$  to denote the total value of rejected jobs. Now, the cost of schedule  $S$  is

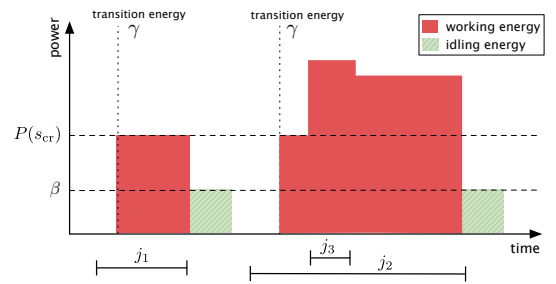
$$\text{cost}(S) := E_{\text{sleep}}^S + E_{\text{idle}}^S + E_{\text{work}}^S + V_{\text{rej}}^S. \quad (1)$$

We seek online strategies yielding a provably good schedule. More formally, we measure the quality of online strategies by their competitive factor: For an online algorithm  $A$  and a problem instance  $I$  let  $A(I)$  denote the resulting schedule and  $O(I)$  an optimal schedule for  $I$ . Then,  $A$  is said to be  $c$ -competitive if  $\sup_I \frac{\text{cost}(A(I))}{\text{cost}(O(I))} \leq c$ .

We define the *system energy*  $E_{\text{sys}}^S$  of a schedule to be the energy needed to hold the system awake (whilst idling and working). That is, if  $S$  is awake for a total of  $x$  time units,  $E_{\text{sys}}^S = x\beta$ . Note that  $E_{\text{sys}}^S \leq E_{\text{idle}}^S + E_{\text{work}}^S$ . The *critical speed* of the power function is defined as  $s_{\text{cr}} := \arg \min_{s \geq 0} P_{\alpha,\beta}(s)/s$  (cf. also [14, 15]). If job  $j$  is processed at constant speed  $s$  its energy usage is  $w_j \cdot P_{\alpha,\beta}(s)/s$ . Thus, assuming that  $j$  is the only job in the system and ignoring its deadline,  $s_{\text{cr}}$  is the energy-optimal speed to process  $j$ . One can easily check that  $s_{\text{cr}}^\alpha = \frac{\beta}{\alpha-1}$ . Given a job  $j$ , let  $\delta_j := v_j/w_j$  denote the job's *value density*. Following [12] and [16], we define the *profitable speed*  $s_{j,p}$  of job  $j$  to be the maximum speed for which its processing may be profitable. More formally,  $s_{j,p} := \max \{s \geq 0 \mid w_j \cdot P_{\alpha,0}(s)/s \leq v_j\}$ . Note that the definition is with respect to  $P_{\alpha,0}$ , i.e., it ignores the system energy. The profitable speed can be more explicitly characterized by  $s_{j,p}^{\alpha-1} = \delta_j$ . It is easy to see that a schedule that processes  $j$  at average speed faster than  $s_{j,p}$  cannot be optimal: rejecting  $j$  and idling during the former execution phase would be more profitable. See Figure 1 for an illustration of these notions.



(a) Our algorithm tries to use job speeds that essentially stay in the shaded interval.



(b) A sample schedule and the involved energy types.

Fig. 1

**Optimal Available & Structural Properties.** One of the first online algorithms for dynamic speed scaling was Optimal Available (OA) due to [17]. As it is an essential building block not only of our but many algorithms for speed scaling, we give a short recap on its idea (see [5] for a thorough discussion and analysis). At any time, OA computes the optimal offline schedule assuming that no further jobs arrive. This optimal offline schedule is computed as follows: Let the density of an interval  $I$  be defined as  $w(I)/|I|$ . Here,  $w(I)$  denotes the total work of jobs  $j$  with  $[r_j, d_j] \subseteq I$  and  $|I|$  the length of  $I$ . Now, whenever a job arrives OA computes so-called *critical intervals* by iteratively choosing an interval of maximum density. Jobs are then scheduled at a speed equal to the density of the corresponding critical interval using the earliest deadline first policy. Let us summarize several structural facts known about the OA schedule.

**Fact 1** *Let  $S$  and  $S'$  denote the OA schedules just before and after  $j$ 's arrival. We use  $S(j)$  and  $S'(j)$  to denote  $j$ 's speed in the corresponding schedule.*

- (a) *The speed function of  $S$  (and  $S'$ ) is a non-increasing staircase function.*
- (b) *The minimal speed of  $S'$  during  $[r_j, d_j]$  is at least  $S'(j)$ .*
- (c) *Let  $I$  be an arbitrary period of time during  $[r_j, d_j]$  (not necessarily an interval). Moreover, let  $W$  denote the total amount of work scheduled by  $S$  and  $W'$  the one scheduled by  $S'$  during  $I$ . Then the inequality  $W \leq W' \leq W + w_j$  holds.*
- (d) *The speed of any  $j' \neq j$  can only increase due to  $j$ 's arrival:  $S'(j') \geq S(j')$ .*

### 3 Lower Bound for Rejection-Oblivious Algorithms

This section considers a class of simple, deterministic online algorithms that we call *rejection-oblivious*. When a job arrives, a rejection-oblivious algorithm decides whether to accept or reject the job. This decision is based solely on the processor's current state (sleeping, idling, working), its current workload, and the job's properties. Especially it does not take former decisions into account. An example for such an algorithm is  $PS(c)$  in [12]. For a suitable parameter  $c$ , it is  $\alpha^\alpha + 2e\alpha$ -competitive in a model without sleep state. In this section we show that in our model (i.e., with a sleep state) no rejection-oblivious algorithm can be competitive. More exactly, the competitiveness of any such algorithm can become arbitrarily large. We identify the jobs' value density as a crucial parameter for the competitiveness of these algorithms.

**Theorem 1.** *The competitiveness of any rejection-oblivious algorithm  $A$  is unbounded. More exactly, for any  $A$  there is a problem instance  $I$  with competitive factor  $\geq \delta_{\max} \frac{s_{\text{cr}}}{P_{\alpha, \beta}(s_{\text{cr}})}$ . Here,  $\delta_{\max}$  is the maximum value density of jobs from  $I$ .*

*Proof.* For  $A$  to be competitive, there must be some  $x \in \mathbb{R}$  such that, while  $A$  is asleep, all jobs of value at most  $x$  are rejected (independent of their work and deadlines). Otherwise, we can define a sequence of  $n$  identical jobs  $1, 2, \dots, n$  of arbitrary small value  $\epsilon$ . W.l.o.g., we release them such that  $A$  goes to sleep during  $[d_{j-1}, r_j]$  (otherwise  $A$  consumes an infinite amount of energy). Thus,  $A$ 's cost is at least  $n\gamma$ . If instead considering schedule  $S$  that rejects all jobs, we have  $\text{cost}(S) = n\epsilon$ . For  $\epsilon \rightarrow 0$  we see that  $A$ 's competitive ratio is unbounded.

So, let  $x \in \mathbb{R}$  be such that  $A$  rejects any job of value at most  $x$  whilst asleep. Consider  $n$  jobs of identical value  $x$  and work  $w$ . For each job, the deadline is set such that  $w = s_{\text{cr}}(d_j - r_j)$ . The jobs are released in immediate succession, i.e.,  $r_j = d_{j-1}$ . Algorithm  $A$  rejects all jobs, incurring cost  $n\gamma$ . Let  $S$  denote the schedule that accepts all jobs and processes them at speed  $s_{\text{cr}}$ . The cost of  $S$  is given by  $\text{cost}(S) = \gamma + nw \frac{P_{\alpha, \beta}(s_{\text{cr}})}{s_{\text{cr}}}$ . Thus,  $A$ 's competitive ratio is at least

$$\frac{n\gamma}{\gamma + nw \frac{P_{\alpha, \beta}(s_{\text{cr}})}{s_{\text{cr}}}} = \delta_{\max} \frac{1}{\frac{\gamma}{nw} + \frac{P_{\alpha, \beta}(s_{\text{cr}})}{s_{\text{cr}}}}.$$

For  $n \rightarrow \infty$  we get the lower bound  $\delta_{\max} \frac{s_{\text{cr}}}{P_{\alpha, \beta}(s_{\text{cr}})}$ .  $\square$

## 4 Algorithm & Analysis

In the following, we use  $A$  to refer to both our algorithm and the schedule it produces; which is meant should be clear from the context. As most algorithms in this area (see, e.g., [4, 6, 12, 14, 15]),  $A$  relies heavily on the good structural properties of OA and its wide applicability to variants of the original energy-oriented scheduling model of Yao, Demers, and Shenker [17]. It essentially consists of two components, the *rejection policy* and the *scheduling policy*. The rejection policy decides which jobs to accept or reject, while the scheduling policy ensures that all accepted jobs are finished until their deadline. Our rejection policy is an extension of the one used by the algorithm PS in [12]. It ensures that we process only jobs that have a reasonable high value (value  $>$  planned energy investment) and that we do not awake from sleep for very cheap jobs. The scheduling policy controls the speed, the job assignment, and the current mode of the processor. It is a straightforward adaption of the algorithm used in [14]. However, its analysis proves to be more involved because we have to take into account its interaction with the rejection policy and that the job sets scheduled by the optimal algorithm and  $A$  may be quite different.

```

1 {at any time  $t$  and for  $x$  equal to current idle cost}
2 on arrival of job  $j$ :
3   {let  $s_{\text{OA}}$  be  $\text{OA}^t$ 's speed for  $j$  if it were accepted}
4   reject if  $\delta_j < s_{\text{cr}}^{\alpha-1}/\alpha c_2^{\alpha-1}$  or  $v_j < c_1 x$  or  $s_{\text{OA}} > c_2 s_{j,p}$ 
5
6 depending on current mode:
7   {let  $\rho_t$  denote  $\text{OA}^t$ 's speed planned for for the current time  $t$ }
8   working:
9     if no remaining work: switch to idle mode
10    otherwise: work at speed  $\max(\rho_t, s_{\text{cr}})$  with earliest deadline first
11  idling:
12    if  $x \geq \gamma$  : switch to sleep mode
13    if  $\rho_t > s_{\text{cr}}$ : switch to work mode
14  sleeping:
15    if  $\rho_t > s_{\text{cr}}$ : switch to work mode

```

Listing 1: Rejection-oblivious online scheduler  $A$ .

The following description assumes a continuous recomputation of the current OA schedule. See Listing 1 for the corresponding pseudocode. It is straightforward to implement  $A$  such that the planned schedule is recomputed only when new jobs arrive.

*Scheduling Policy.* All accepted jobs are scheduled according to the earliest deadline first rule. At any time, the processor speed is computed based on the OA schedule. Use  $\text{OA}^t$  to denote the schedule produced by OA if given the remaining (accepted) work at time  $t$  and the power function  $P_{\alpha,0}$ . Let  $\rho_t$  denote the speed planned by  $\text{OA}^t$  at time  $t$ .  $A$  puts the processor either in *working*, *idling*, or *sleeping* mode. During working mode the processor speed is set to  $\max(\rho_t, s_{\text{cr}})$  until there is no more remaining work. Then, speed is set to zero and the processor starts idling. When idling or sleeping, we switch to the working mode only when  $\rho_t$  becomes larger than  $s_{\text{cr}}$ . When the amount of energy spent in the current idle interval equals the transition energy  $\gamma$  (i.e., after time  $\gamma/P_{\alpha,\beta}(0)$ ) the processor is suspended to sleep.

*Rejection Policy.* Let  $c_1$  and  $c_2$  be parameters to be determined later. Consider the arrival of a new job  $j$  at time  $r_j$ . Reject it immediately if  $\delta_j < s_{cr}^{\alpha-1}/\alpha c_2^{\alpha-1}$ . Otherwise, define the *current idle cost*  $x \in [0, \gamma]$  depending on the processor's state as follows: (i) zero if it is working, (ii) the length of the current idle interval times  $\beta$  if it is idle, and (iii)  $\gamma$  if it is asleep. If  $v_j < c_1 x$ , the job is rejected. Otherwise, compute the job's speed  $s_{OA}$  which would be assigned by  $OA^{r_j}$  if it were accepted. Reject the job if  $s_{OA} > c_2 s_{j,p}$ , accept otherwise.

#### 4.1 Bounding the Different Portions of the Cost

In the following, let  $O$  denote an optimal schedule. Remember that  $\text{cost}(A) = E_{\text{sleep}}^A + E_{\text{idle}}^A + E_{\text{work}}^A + V_{\text{rej}}^A$ . We bound each of the three terms  $E_{\text{sleep}}^A + E_{\text{idle}}^A$ ,  $E_{\text{work}}^A$ , and  $V_{\text{rej}}^A$  separately in Lemma 1, Lemma 2, and Lemma 3, respectively. Eventually, Section 4.2 combines these bounds and yields our main result: a nearly tight competitive factor depending on the maximum value density of the problem instance.

**Lemma 1 (Sleep and Idle Energy).**  $E_{\text{sleep}}^A + E_{\text{idle}}^A \leq 6E_{\text{sleep}}^O + 2E_{\text{sys}}^O + \frac{4}{c_1} V_{\text{rej}}^O$

*Proof.* Let us first consider  $E_{\text{idle}}^A$ . Partition the set of idle intervals under schedule  $A$  into three disjoint subsets  $\mathcal{I}_1$ ,  $\mathcal{I}_2$ , and  $\mathcal{I}_3$  as follows:

- $\mathcal{I}_1$  contains idle intervals not intersecting any sleep interval of  $O$ . By definition, the total length of idle intervals from  $\mathcal{I}_1$  is bounded by the time  $O$  is awake. Thus, the total cost of  $\mathcal{I}_1$  is at most  $E_{\text{sys}}^O$ .
- For each sleep interval  $I$  of  $O$ ,  $\mathcal{I}_2$  contains any idle interval  $X$  that is not the last idle interval having a nonempty intersection with  $I$  and that is completely contained within  $I$  (note that the former requirement is redundant if the last intersecting idle interval is not completely contained in  $I$ ). Consider any  $X \in \mathcal{I}_2$  intersecting  $I$  and let  $j$  denote the first job processed by  $A$  after  $X$ . It is easy to see that we must have  $[r_j, d_j] \subseteq I$ . Thus,  $O$  has rejected  $j$ . But since  $A$  accepted  $j$ , we must have  $v_j \geq c_1 |X| \beta$ . This implies that the total cost of  $\mathcal{I}_2$  cannot exceed  $V_{\text{rej}}^O/c_1$ .
- $\mathcal{I}_3$  contains all remaining idle intervals. By definition, the first sleep interval of  $O$  can intersect at most one such idle interval, while the remaining sleep intervals of  $O$  can be intersected by at most two such idle intervals. Thus, if  $m$  denotes the number of sleep intervals under schedule  $O$ , we get  $|\mathcal{I}_3| \leq 2m - 1$ . Our sleeping strategy ensures that the cost of each single idle interval is at most  $\gamma$ . Using this and the definition of sleeping energy, the total cost of  $\mathcal{I}_3$  is upper bounded by  $(2m - 1)\gamma = 2E_{\text{sleep}}^O + \gamma$ .

Together, we get  $E_{\text{idle}}^A \leq E_{\text{sys}}^O + V_{\text{rej}}^O/c_1 + 2E_{\text{sleep}}^O + \gamma$ . Moreover, without loss of generality we can bound  $\gamma$  by  $V_{\text{rej}}^O/c_1 + E_{\text{sleep}}^O$ : if not both  $A$  and  $O$  reject all incoming jobs (in which case  $A$  would be optimal),  $O$  will either accept at least one job and thus wake up ( $\gamma \leq E_{\text{sleep}}^O$ ) or reject the first job  $A$  accepts ( $\gamma \leq V_{\text{rej}}^O/c_1$ ). This yields  $E_{\text{idle}}^A \leq E_{\text{sys}}^O + 2V_{\text{rej}}^O/c_1 + 3E_{\text{sleep}}^O$ . For  $E_{\text{sleep}}^A$ , note that any but the first of  $A$ 's sleep intervals is preceded by an idle interval of length  $\gamma/P_{\alpha,\beta}(0)$ . Each such idle interval has cost  $\gamma$ , so we get  $E_{\text{sleep}}^A \leq E_{\text{idle}}^A$ . The lemma's statement follows by combining the bounds for  $E_{\text{idle}}^A$  and  $E_{\text{sleep}}^A$ .  $\square$

**Lemma 2 (Working Energy).**  $E_{\text{work}}^A \leq \alpha^\alpha E_{\text{work}}^O + c_2^{\alpha-1} \alpha^2 V_{\text{rej}}^O$

The proof of Lemma 2 is based on the standard amortized local competitiveness argument, first used by Bansal, Kimbrel, and Pruhs [5]. Although technically quite similar to the typical

argumentation, our proof must carefully consider the more complicated rejection policy (compared to [12]), while simultaneously handle the different processor states.

Given a schedule  $S$ , let  $E_{\text{work}}^S(t)$  denote the working energy spent until time  $t$  and  $V_{\text{rej}}^S(t)$  the discarded value until time  $t$ . We show that at any time  $t \in \mathbb{R}_{\geq 0}$  the *amortized energy inequality*

$$E_{\text{work}}^A(t) + \Phi(t) \leq \alpha^\alpha E_{\text{work}}^O(t) + c_2^{\alpha-1} \alpha^2 V_{\text{rej}}^O(t) \quad (2)$$

holds. Here,  $\Phi$  is a potential function to be defined in a suitable way. It is constructed such that the following conditions hold:

- (i) *Boundary Condition:* At the beginning and end we have  $\Phi(t) = 0$ .
- (ii) *Running Condition:* At any time  $t$  when no job arrives we have

$$\frac{dE_{\text{work}}^A(t)}{dt} + \frac{d\Phi(t)}{dt} \leq \alpha^\alpha \frac{dE_{\text{work}}^O(t)}{dt} + c_2^{\alpha-1} \alpha^2 \frac{dV_{\text{rej}}^O(t)}{dt}. \quad (3)$$

- (iii) *Arrival Condition:* At any time  $t$  when a job arrives we have

$$\Delta E_{\text{work}}^A(t) + \Delta \Phi(t) \leq \alpha^\alpha \Delta E_{\text{work}}^O(t) + c_2^{\alpha-1} \alpha^2 \Delta V_{\text{rej}}^O(t). \quad (4)$$

The  $\Delta$ -terms denote the corresponding change caused by the job arrival.

Once these are proven, amortized energy inequality follows by induction: It obviously holds for  $t = 0$ , and Conditions (ii) and (iii) ensure that it is never violated. Applying Condition (i) yields Lemma 2. The crucial part is to define a suitable potential function  $\Phi$ . Our analysis combines aspects from both [12] and [14]. Different rejection decisions of our algorithm  $A$  and the optimal algorithm  $O$  require us to handle possibly different job sets in the analysis, while the sleep management calls for a careful handling of the processor's current state.

*Construction of  $\Phi$ .* Consider an arbitrary time  $t \in \mathbb{R}_{\geq 0}$ . Let  $w_t^A(t_1, t_2)$  denote the remaining work at time  $t$  accepted by schedule  $A$  with deadline in  $(t_1, t_2]$ . We call the expression  $\frac{w_t^A(t_1, t_2)}{t_2 - t_1}$  the *density* of the interval  $(t_1, t_2]$ . Next, we define *critical intervals*  $(\tau_{i-1}, \tau_i]$ . For this purpose, set  $\tau_0 := t$  and define  $\tau_i$  iteratively to be the maximum time that maximizes the density  $\rho_i := \frac{w_t^A(\tau_{i-1}, \tau_i)}{\tau_i - \tau_{i-1}}$  of the interval  $(\tau_{i-1}, \tau_i]$ . We end at the first index  $l$  with  $\rho_l \leq s_{\text{cr}}$  and set  $\tau_l = \infty$  and  $\rho_l = s_{\text{cr}}$ . Note that  $\rho_1 > \rho_2 > \dots > \rho_l = s_{\text{cr}}$ . Now, for a schedule  $S$  let  $w_t^S(i)$  denote the remaining work at time  $t$  with deadline in the  $i$ -th critical interval  $(\tau_{i-1}, \tau_i]$  accepted by schedule  $S$ . The potential function is defined as  $\Phi(t) := \alpha \sum_{i=1}^l \rho_i^{\alpha-1} (w_t^A(i) - \alpha w_t^O(i))$ . It quantifies how far  $A$  is ahead or behind in terms of energy. The densities  $\rho_i$  essentially correspond to OA's speed levels, but are adjusted to  $A$ 's usage of OA. Note that whenever  $A$  is in working mode its speed equals  $\rho_1 \geq s_{\text{cr}}$ .

It remains to prove the boundary, running, and arrival conditions. The boundary condition is trivially true as both  $A$  and  $O$  have no remaining work at the beginning and end. For the running and arrival conditions, see Propositions 1 and 2, respectively.

**Proposition 1.** *The running condition holds. That is, at any time  $t$  when no job arrives we have*

$$\frac{dE_{\text{work}}^A(t)}{dt} + \frac{d\Phi(t)}{dt} \leq \alpha^\alpha \frac{dE_{\text{work}}^O(t)}{dt} + c_2^{\alpha-1} \alpha^2 \frac{dV_{\text{rej}}^O(t)}{dt}.$$

*Proof.* Because no jobs arrive we have  $\frac{dV_{\text{rej}}^O(t)}{dt} = 0$ . Let  $s_A$  denote the speed of  $A$  and  $s_O$  the speed of  $O$ . Depending on these speeds, we distinguish four cases:



**Case 1:**  $s_O = 0, s_A = 0$

In this case  $\frac{dE_{\text{work}}^A(t)}{dt} = \frac{dE_{\text{work}}^O(t)}{dt} = \frac{d\Phi(t)}{dt} = 0$ . Thus, the Running Condition (3) holds.

**Case 2:**  $s_O = 0, s_A > 0$

Since  $s_A > 0$ , algorithm  $A$  is in working mode and we have  $s_A = \rho_1 \geq s_{\text{cr}}$ . Moreover,  $\frac{dE_{\text{work}}^A(t)}{dt} = P_{\alpha,\beta}(s_A)$ ,  $\frac{d\Phi(t)}{dt} = -\alpha s_A^\alpha$ , and  $\frac{dE_{\text{work}}^O(t)}{dt} = 0$ . We get

$$\begin{aligned} \frac{dE_{\text{work}}^A(t)}{dt} + \frac{d\Phi(t)}{dt} - \alpha^\alpha \frac{dE_{\text{work}}^O(t)}{dt} &= P_{\alpha,\beta}(s_A) - \alpha s_A^\alpha \\ &= \beta - (\alpha - 1)s_A^\alpha \leq \beta - (\alpha - 1)s_{\text{cr}}^\alpha = 0. \end{aligned}$$

**Case 3:**  $s_O > 0, s_A = 0$

In this case  $l = 1$  and thus  $\rho_1 = s_{\text{cr}}$ . The terms in Inequality (3) become  $\frac{dE_{\text{work}}^A(t)}{dt} = 0$ ,  $\frac{d\Phi(t)}{dt} = \alpha^2 s_{\text{cr}}^{\alpha-1} s_O$ , and  $\frac{dE_{\text{work}}^O(t)}{dt} = P_{\alpha,\beta}(s_O)$ . We get

$$\begin{aligned} \frac{dE_{\text{work}}^A(t)}{dt} + \frac{d\Phi(t)}{dt} - \alpha^\alpha \frac{dE_{\text{work}}^O(t)}{dt} &= \alpha^2 s_{\text{cr}}^{\alpha-1} s_O - \alpha^\alpha P_{\alpha,\beta}(s_O) \\ &\leq \alpha^2 s_{\text{cr}}^{\alpha-1} s_O - \alpha^\alpha s_O \frac{P_{\alpha,\beta}(s_{\text{cr}})}{s_{\text{cr}}} \leq s_O s_{\text{cr}}^{\alpha-1} (\alpha^2 - \alpha^\alpha) \leq 0. \end{aligned}$$

**Case 4:**  $s_O > 0, s_A > 0$

Because of  $s_A > 0$  we know  $A$  is in the working state and, thus,  $s_A = \rho_1 \geq s_{\text{cr}}$ . So, this time we have  $\frac{dE_{\text{work}}^A(t)}{dt} = P_{\alpha,\beta}(s_A)$ ,  $\frac{d\Phi(t)}{dt} = -\alpha s_A^\alpha + \alpha^2 \rho_k^{\alpha-1} s_O$ , and  $\frac{dE_{\text{work}}^O(t)}{dt} = P_{\alpha,\beta}(s_O)$ . We get

$$\begin{aligned} \frac{dE_{\text{work}}^A(t)}{dt} + \frac{d\Phi(t)}{dt} - \alpha^\alpha \frac{dE_{\text{work}}^O(t)}{dt} &= P_{\alpha,\beta}(s_A) - \alpha s_A^\alpha + \alpha^2 \rho_k^{\alpha-1} s_O - \alpha^\alpha P_{\alpha,\beta}(s_O) \\ &\leq s_A^\alpha - \alpha s_A^\alpha + \alpha^2 s_A^{\alpha-1} s_O - \alpha^\alpha s_O^\alpha \leq 0. \end{aligned}$$

The last inequality follows from the same argument as in [5]: Divide by  $s_O^\alpha$  and substitute  $z = s_A/s_O$ . It becomes equivalent to  $(1 - \alpha)z^\alpha + \alpha^2 z^{\alpha-1} - \alpha^\alpha \leq 0$ . Differentiating with respect to  $z$  yields the correctness.  $\square$

**Proposition 2.** *The arrival condition holds. That is, at any time  $t$  when a job arrives we have*

$$\Delta E_{\text{work}}^A(t) + \Delta\Phi(t) \leq \alpha^\alpha \Delta E_{\text{work}}^O(t) + c_2^{\alpha-1} \alpha^2 \Delta V_{\text{rej}}^O(t).$$

Here, the  $\Delta$ -terms denote the corresponding change caused by the job arrival.

*Proof.* The arrival of a job  $j$  does not change the energy invested so far, thus  $\Delta E_{\text{work}}^A(t) = 0$  and  $\Delta E_{\text{work}}^O(t) = 0$ . If  $A$  rejects  $j$ , we have  $\Delta\Phi(t) \leq 0$  and  $\Delta V_{\text{rej}}^O(t) \geq 0$ , thus the Arrival Condition (4) holds. So assume  $A$  accepts  $j$ . The arrival of  $j$  may change the critical intervals and their densities significantly. However, as pointed out in [5], these changes can be broken down into a series of simpler changes affecting at most two adjacent critical intervals. Thus, we first consider the effect of arrivals which do not change the critical intervals. Afterward, we use the technique from [5] to reduce an arbitrary change to these simple cases.

**Case 1:** The critical intervals remain unchanged and only  $\rho_k$  for  $k < l$  changes.

Let  $\rho_k$  and  $\rho'_k$  denote the densities of  $(\tau_{k-1}, \tau_k]$  just before and after  $j$ 's arrival, respectively. That is,  $\rho'_k = \rho_k + \frac{w_j}{\tau_k - \tau_{k-1}}$ . Note that  $\rho'_k$  is the speed planned by  $\text{OA}^t$  for job  $j$ . Because  $A$  accepted  $j$  we have  $\rho'_k \leq c_2 s_{j,p}$ . If  $j$  is rejected by  $O$ , we have  $\Delta V_{\text{rej}}^O(t) = v_j$ . Since only the  $k$ -th critical interval is affected, the change in the potential function is given by

$$\Delta\Phi(t) = \alpha \rho_k'^{\alpha-1} (w_t^A(k) + w_j - \alpha w_t^O(k)) - \alpha \rho_k^{\alpha-1} (w_t^A(k) - \alpha w_t^O(k)).$$

Now, we compute, analogously to Lemma 4 in [12], that  $\Delta\Phi(t)$  equals

$$\begin{aligned} & \alpha\rho_k^{\alpha-1} (w_t^A(k) + w_j - \alpha w_t^O(k)) - \alpha\rho_k^{\alpha-1} (w_t^A(k) - \alpha w_t^O(k)) \\ & \leq \alpha\rho_k^{\alpha-1} (w_t^A(k) + w_j) - \alpha\rho_k^{\alpha-1} w_t^A(k) = \alpha \frac{(w_t^A(k) + w_j)^\alpha - w_t^A(k)^\alpha}{(\tau_k - \tau_{k-1})^{\alpha-1}} \\ & \leq \alpha^2 \frac{(w_t^A(k) + w_j)^{\alpha-1} w_j}{(\tau_k - \tau_{k-1})^{\alpha-1}} = \alpha^2 \rho_k^{\alpha-1} w_j \leq \alpha^2 (c_2 s_{j,p})^{\alpha-1} w_j = c_2^{\alpha-1} \alpha^2 v_j. \end{aligned}$$

The penultimate inequality uses the fact that  $f(x) = x^\alpha$  is convex, yielding  $f(y) - f(x) \leq f'(y) \cdot (y - x)$  for all  $y > x$ . This implies the Arrival Condition (4). If  $j$  is accepted by  $O$ , we have  $\Delta V_{\text{rej}}^O(t) = 0$  and  $\Delta\Phi(t)$  becomes

$$\alpha\rho_k^{\alpha-1} (w_t^A(k) + w_j - \alpha (w_t^O(k) + w_j)) - \alpha\rho_k^{\alpha-1} (w_t^A(k) - \alpha w_t^O(k)).$$

In the same way as in [5], we now get  $\Delta\Phi(t) \leq 0$ .

**Case 2:** Only the amount of work assigned to the last critical interval increases.

Remember that, by definition,  $\tau_l = \infty$  and  $\rho_l = s_{\text{cr}}$ . Since  $j$  is accepted by  $A$  we have  $s_{\text{cr}}^{\alpha-1} \leq \alpha c_2^{\alpha-1} \delta_j$ . If  $O$  rejects  $j$  we have  $\Delta V_{\text{rej}}^O(t) = v_j$  and the Arrival Condition (4) follows from

$$\begin{aligned} \Delta\Phi(t) &= \alpha\rho_l^{\alpha-1} (w_t^A(l) + w_j - \alpha w_t^O(l)) - \alpha\rho_l^{\alpha-1} (w_t^A(l) - \alpha w_t^O(l)) \\ &= \alpha s_{\text{cr}}^{\alpha-1} w_j \leq \alpha^2 c_2^{\alpha-1} \delta_j w_j = c_2^{\alpha-1} \alpha^2 v_j. \end{aligned}$$

If  $O$  accepts  $j$ ,  $\Delta V_{\text{rej}}^O(t) = 0$  and the Arrival Condition (4) is implied by

$$\begin{aligned} \Delta\Phi(t) &= \alpha\rho_l^{\alpha-1} (w_t^A(l) + w_j - \alpha (w_t^O(l) + w_j)) - \alpha\rho_l^{\alpha-1} (w_t^A(l) - \alpha w_t^O(l)) \\ &= \alpha\rho_l^{\alpha-1} w_j (1 - \alpha) \leq 0. \end{aligned}$$

Let us now consider the arrival of an arbitrary job  $j$ . The idea is to split this job into two jobs  $j_1$  and  $j_2$  with the same release time, deadline, and value density as  $j$ . Their total work equals  $w_j$ . Let  $x$  denote the size of  $j_1$ . We determine a suitable  $x$  by continuously increasing  $x$  from 0 to  $w_j$  until two critical intervals merge or one critical interval splits. The arrival of  $j_1$  can then be handled by one of the above cases, while  $j_2$  is treated recursively in the same way as  $j$ . For details, see [5] or [14].  $\square$

**Bounding the Rejected Value.** In the following we bound the total value  $V_{\text{rej}}^A$  of jobs rejected by  $A$ . The general idea is similar to the one by Chan, Lam, and Li [12]. However, in contrast to the simpler model without sleep states, we must handle small-valued jobs of high density explicitly (cf. Section 3). Moreover, the sleeping policy introduces an additional difficulty: our algorithm does not preserve all structural properties of an OA schedule (cf. Fact 1). This prohibits a direct mapping between the energy consumption of algorithm  $A$  and of the intermediate OA schedules during a *fixed time interval*, as used in the corresponding proof in [12]. Indeed, the actual energy used by  $A$  during a fixed time interval may decrease compared to the energy planned by the intermediate OA schedule, as  $A$  may decide to raise the speed to  $s_{\text{cr}}$  at certain points in the schedule. Thus, to bound the value of a job rejected by  $A$  but processed by the optimal algorithm for a relatively long time, we have to consider the energy usage for the workload OA planned for that time (instead of the actual energy usage for the workload  $A$  processed during that time, which might be quite different).

**Lemma 3 (Rejected Value).** *Let  $\delta_{\max}$  be the maximum value density of jobs of value less than  $c_1\gamma$  and consider an arbitrary parameter  $b \geq 1/c_2$ . Then,  $A$ 's rejected value is at most*

$$V_{\text{rej}}^A \leq \max\left(\delta_{\max} \frac{s_{\text{cr}}}{P_{\alpha,\beta}(s_{\text{cr}})}, b^{\alpha-1}\right) E_{\text{work}}^O + \frac{b^{\alpha-1}}{(c_2b-1)^\alpha} E_{\text{work}}^A + V_{\text{rej}}^O.$$

*Proof.* Partition the jobs rejected by  $A$  into two disjoint subsets  $J_1$  (jobs rejected by both  $A$  and  $O$ ) and  $J_2$  (jobs rejected by  $A$  only). The total value of jobs in  $J_1$  is at most  $V_{\text{rej}}^O$ . Thus, it suffices to show that the total value of  $J_2$  is bounded by

$$\max\left(\delta_{\max} \frac{s_{\text{cr}}}{P_{\alpha,\beta}(s_{\text{cr}})}, b^{\alpha-1}\right) E_{\text{work}}^O + \frac{b^{\alpha-1}}{(c_2b-1)^\alpha} E_{\text{work}}^A.$$

To this end, let  $j \in J_2$ . Remember that, because of the convexity of the power function,  $O$  can be assumed to process  $j$  at a constant speed  $s_O$ . Otherwise processing  $j$  at its average speed could only improve the schedule. Let us distinguish three cases, depending on the reason for which  $A$  rejected  $j$ :

**Case 1:**  $j$  got rejected because of  $\delta_j < \frac{s_{\text{cr}}^{\alpha-1}}{\alpha c_2^{\alpha-1}}$ .

Let  $E_{\text{work}}^O(j)$  denote the working energy invested by  $O$  into job  $j$ . Using the rejection condition we can compute

$$E_{\text{work}}^O(j) = \frac{P_{\alpha,\beta}(s_O)}{s_O} w_j \geq \frac{P_{\alpha,\beta}(s_{\text{cr}})}{s_{\text{cr}}} w_j \geq s_{\text{cr}}^{\alpha-1} w_j > \alpha c_2^{\alpha-1} v_j.$$

Together with  $b \geq 1/c_2$  we get  $v_j < b^{\alpha-1} E_{\text{work}}^O(j)$ .

**Case 2:**  $j$  got rejected because of  $v_j < c_1x$

As in the algorithm description, let  $x \in [0, \gamma]$  denote the current idle cost at time  $r_j$ . Since  $j$ 's value is less than  $c_1x \leq c_1\gamma$ , we have  $\delta_j \leq \delta_{\max}$ . We get

$$E_{\text{work}}^O(j) = \frac{P_{\alpha,\beta}(s_O)}{s_O} w_j = \frac{P_{\alpha,\beta}(s_O)}{s_O} \frac{v_j}{\delta_j} \geq \frac{P_{\alpha,\beta}(s_{\text{cr}})}{s_{\text{cr}}} \frac{v_j}{\delta_{\max}},$$

which eventually yields  $v_j \leq \delta_{\max} \frac{s_{\text{cr}}}{P_{\alpha,\beta}(s_{\text{cr}})} E_{\text{work}}^O(j)$ .

**Case 3:**  $j$  got rejected because of  $s_{\text{OA}} > c_2s_{j,p}$

Here,  $s_{\text{OA}}$  denotes the speed  $\text{OA}^{r_j}$  would assign to  $j$  if it were accepted. We use  $\text{OA}_-^{r_j}$  to refer to the  $\text{OA}$  schedule at time  $r_j$  without  $j$ . Let  $b_j := s_{j,p}/s_O$ . We bound  $v_j$  in different ways, depending on  $b_j$ . If  $b_j$  is small (i.e.,  $b_j \leq b$ ) we use

$$E_{\text{work}}^O(j) \geq \frac{P_{\alpha,0}(s_O)}{s_O} w_j = \frac{P_{\alpha,0}(s_{j,p}/b_j)}{s_{j,p}/b_j} w_j = \frac{s_{j,p}^{\alpha-1}}{b_j^{\alpha-1}} w_j = \frac{v_j}{b_j^{\alpha-1}}.$$

That is, we have  $v_j \leq b_j^{\alpha-1} E_{\text{work}}^O(j)$ . Otherwise, if  $b_j$  is relatively large,  $v_j$  is bounded by  $E_{\text{work}}^A$ . Let  $I$  denote the period of time when  $O$  processes  $j$  at constant speed  $s_O$  and let  $W$  denote the work processed by  $\text{OA}_-^{r_j}$  during this time. Since  $I \subseteq [r_j, d_j]$ , Fact 1((b)) implies that  $\text{OA}^{r_j}$ 's speed during  $I$  is at least  $s_{\text{OA}} > c_2s_{j,p}$ . Thus, the total amount of work processed by  $\text{OA}^{r_j}$  during  $I$  is larger than  $c_2s_{j,p} |I|$ . But then, by applying Fact 1((c)), we see that  $W$  must be larger than  $c_2s_{j,p} |I| - w_j$ . Now,  $W$  is a subset of the work processed by  $A$ . Moreover, Fact 1((d)) and the definition of algorithm  $A$  ensure that the speeds used for this work in schedule  $A$  cannot be smaller than the ones used in  $\text{OA}_-^{r_j}$ . Especially, the average speed  $s_{\text{OA}}$  used for this work in schedule  $A$  is at least  $W/|I|$  (the average speed used by  $\text{OA}_-^{r_j}$  for this

work). Let  $E_{\text{work}}^A(W)$  denote the energy invested by schedule  $A$  into the work  $W$ . Then, by exploiting the convexity of the power function, we get

$$\begin{aligned} E_{\text{work}}^A(W) &\geq \frac{P_{\alpha,\beta}(s_{\emptyset})}{s_{\emptyset}} W \geq \frac{P_{\alpha,0}(s_{\emptyset})}{s_{\emptyset}} W = s_{\emptyset}^{\alpha-1} W \geq \frac{W^{\alpha-1}}{|I|^{\alpha-1}} W = |I| \frac{W^{\alpha}}{|I|^{\alpha}} \\ &> |I| (c_2 s_{j,p} - s_O)^{\alpha} = \frac{w_j}{s_O} s_O^{\alpha} (c_2 b_j - 1)^{\alpha} = \frac{(c_2 b_j - 1)^{\alpha}}{b_j^{\alpha-1}} v_j. \end{aligned}$$

That is, we have  $v_j < \frac{b_j^{\alpha-1}}{(c_2 b_j - 1)^{\alpha}} E_{\text{work}}^A(W)$ . Now, let us specify how to choose from these two bounds:

- If  $b_j \leq b$ , we apply the first bound:  $v_j = b_j^{\alpha-1} E_{\text{work}}^O(j) \leq b^{\alpha-1} E_{\text{work}}^O(j)$ .
- Otherwise we have  $b_j > b \geq 1/c_2$ . Note that for  $x > 1/c$  the function  $f(x) = \frac{x^{\alpha-1}}{(cx-1)^{\alpha}}$  decreases. Thus, we get  $v_j < \frac{b^{\alpha-1}}{(c_2 b - 1)^{\alpha}} E_{\text{work}}^A(W)$ .

By combining these cases we get

$$v_j \leq \max\left(\delta_{\max} \frac{s_{cr}}{P_{\alpha,\beta}(s_{cr})}, b^{\alpha-1}\right) E_{\text{work}}^O(j) + \frac{b^{\alpha-1}}{(c_2 b - 1)^{\alpha}} E_{\text{work}}^A(W).$$

Note that both energies referred to,  $E_{\text{work}}^O(j)$  as well as  $E_{\text{work}}^A(W)$ , are mutually different for different jobs  $j$ . Thus, we can combine these inequalities for all jobs  $j \in J_2$  to get the desired result.  $\square$

## 4.2 Putting it All Together.

The following theorem combines the results of Lemma 1, Lemma 2, and Lemma 3.

**Theorem 2.** *Let  $\alpha \geq 2$  and let  $\delta_{\max}$  be the maximum value density of jobs of value less than  $c_1 \gamma$ . Moreover, define  $\eta := \max\left(\delta_{\max} \frac{s_{cr}}{P_{\alpha,\beta}(s_{cr})}, b^{\alpha-1}\right)$  and  $\mu := \frac{b^{\alpha-1}}{(c_2 b - 1)^{\alpha}}$  for a parameter  $b \geq 1/c_2$ . Then,  $A$ 's competitive factor is at most*

$$\max(c_2^{\alpha-1} \alpha^2, \alpha^{\alpha}) (1 + \mu) + \max(2 + \eta, 1 + 4/c_1).$$

*Proof.* Lemma 1 together with the relation  $E_{\text{sys}}^O \leq E_{\text{idle}}^O + E_{\text{work}}^O$  bounds the sleep and idle energy of  $A$  with respect to  $O$ 's cost as  $E_{\text{sleep}}^A + E_{\text{idle}}^A \leq 6E_{\text{sleep}}^O + 2E_{\text{idle}}^O + 2E_{\text{work}}^O + \frac{4}{c_1} V_{\text{rej}}^O$ . For the working energy, Lemma 2 yields  $E_{\text{work}}^A \leq \alpha^{\alpha} E_{\text{work}}^O + c_2^{\alpha-1} \alpha^2 V_{\text{rej}}^O$ . To bound the total value rejected by  $A$  with respect to the cost of  $O$ , we apply Lemma 2 to Lemma 3 and get

$$V_{\text{rej}}^A \leq \eta E_{\text{work}}^O + \mu E_{\text{work}}^A + V_{\text{rej}}^O \leq (\eta + \alpha^{\alpha} \mu) E_{\text{work}}^O + (c_2^{\alpha-1} \alpha^2 \mu + 1) V_{\text{rej}}^O.$$

Using these inequalities, we can bound the cost of  $A$  as follows:

$$\begin{aligned} \text{cost}(A) &\leq 6E_{\text{sleep}}^O + 2E_{\text{idle}}^O + (\alpha^{\alpha} + \alpha^{\alpha} \mu + 2 + \eta) E_{\text{work}}^O \\ &\quad + (c_2^{\alpha-1} \alpha^2 + c_2^{\alpha-1} \alpha^2 \mu + 1 + 4/c_1) V_{\text{rej}}^O. \end{aligned}$$

Since  $6 \leq \alpha^{\alpha} + 2$  for  $\alpha \geq 2$ , we get the following bound on  $A$ 's competitive factor:

$$\frac{\text{cost}(A)}{\text{cost}(O)} \leq \max(c_2^{\alpha-1} \alpha^2, \alpha^{\alpha}) (1 + \mu) + \max(2 + \eta, 1 + 4/c_1).$$

$\square$

By a careful choice of parameters we get a constant competitive ratio if restricting the value density of small-valued jobs accordingly. So, let  $\alpha \geq 2$  and set  $c_2 = \alpha^{\frac{\alpha-2}{\alpha-1}}$ ,  $b = \frac{\alpha+1}{c_2}$ , and  $c_1 = \frac{4}{1+b^{\alpha-1}} \leq 1$ . Applying Theorem 2 using these parameters yields the following results:

**Corollary 1.** *Algorithm A is  $\alpha^\alpha + 2e\alpha + \delta_{\max} \frac{s_{cr}}{P_{\alpha,\beta}(s_{cr})}$ -competitive.*

**Corollary 2.** *Algorithm A is  $\alpha^\alpha + 2e\alpha$ -competitive if we restrict it to instances of maximum value density  $\delta_{\max} := b^{\alpha-1} \frac{P_{\alpha,\beta}(s_{cr})}{s_{cr}}$ . This competitive factor still holds if the restriction is only applied to jobs of value less than  $\frac{4}{1+b^{\alpha-1}}\gamma$ .*

*Proof.* First note the identity  $b^{\alpha-1} = \alpha(1 + 1/\alpha)^{\alpha-1}$ . Moreover, using the definitions from Theorem 2, we see that  $\eta = b^{\alpha-1}$  and  $\alpha^\alpha \mu = b^{\alpha-1}$ . By applying Theorem 2 to our choice of parameters, the competitive factor of A becomes

$$\begin{aligned} \alpha^\alpha(1 + \mu) + 2 + \eta &= \alpha^\alpha + 2 + 2b^{\alpha-1} = \alpha^\alpha + 2(1 + \alpha(1 + 1/\alpha)^{\alpha-1}) \\ &\leq \alpha^\alpha + 2\alpha(1 + 1/\alpha)^\alpha \leq \alpha^\alpha + 2e\alpha. \end{aligned}$$

□

**Corollary 3.** *If only considering instances for which the job values are at least  $\frac{8}{2+3\alpha}\gamma \leq \gamma$ , A's competitive factor is at most  $\alpha^\alpha + 2e\alpha$ .*

*Proof.* Follows from Corollary 2 by using that for  $\alpha \geq 2$  we have  $\frac{4}{1+b^{\alpha-1}} = \frac{4}{1+\alpha(1+1/\alpha)^{\alpha-1}} \leq \frac{4}{1+\frac{3}{2}\alpha} = \frac{8}{2+3\alpha}$ . □

Note that the bound from Corollary 1 is nearly tight with respect to  $\delta_{\max}$  and the lower bound from Theorem 1.

## 5 The Speed-Bounded Case

As stated earlier, our model can be considered as a generalization of [12]. It adds sleep states, leading to several structural difficulties which we solved in the previous section. A further, natural generalization to the model is to cap the speed at some maximum speed  $T$ . Algorithms based on OA often lend themselves to such bounded speed models. In many cases, a canonical adaptation – possibly mixed with a more involved job selection rule – leads to an algorithm for the speed bounded case with similar properties (see, e.g., [7, 12–14]). A notable property of the profit-oriented scheduling model of [12] is that limiting the maximum speed leads to a non-constant competitive factor. Instead, it becomes highly dependent on a job's penalty ratio defined as  $\Gamma_j := s_{j,p}/T$ . They derive a lower bound of  $\Omega(\max(e^{\alpha-1}/\alpha, \Gamma^{\alpha-2+1/\alpha}))$  where  $\Gamma = \max \Gamma_j$ . Since our model generalizes their model, this bound transfers immediately to our setting (for the case  $\beta = \gamma = 0$ ). On the positive side we can adapt our algorithm, similar to [12], by additionally rejecting a job if its speed planned by OA is larger than  $T$  (cf. rejection condition in algorithm description, Section 4). Our main theorem from Section 4 becomes

**Theorem 3.** *Let  $\alpha \geq 2$  and let  $\delta_{\max}$  be the maximum value density of jobs of value less than  $c_1\gamma$ . Moreover, define  $\eta := \max(\delta_{\max} \frac{s_{cr}}{P_{\alpha,\beta}(s_{cr})}, \Gamma^{\alpha-1} b^{\alpha-1})$  and  $\mu := \Gamma^{\alpha-1} \frac{b^{\alpha-1}}{(b-1)^\alpha}$  for  $b \geq 1$ . Then, A's competitive factor is at most*

$$\alpha^\alpha(1 + \mu) + \max(2 + \eta, 1 + 4/c_1).$$

*Proof (sketch).* Note that the results from Lemmas 1 and 2 remain valid without any changes, as an additional rejection rule does not influence the corresponding proofs. The only lemma affected by the changed algorithm is Lemma 3. In its proof, we have to consider an additional rejection case, namely that job  $j$  got rejected because of  $s_{OA} > T = \frac{1}{\Gamma_j} s_{j,p}$ . This can be

handled completely analogously to Case 3 in the proof, using the factor  $\frac{1}{\Gamma_j}$  instead of  $c_2$ . We get the bounds  $v_j \leq b_j^{\alpha-1} E_{\text{work}}^O(j)$  and  $v_j < b_j^{\alpha-1} / (b_j/\Gamma_j - 1)^\alpha E_{\text{work}}^A(W)$ . If  $b_j \leq \Gamma_j b$  this yields  $v_j \leq \Gamma_j^{\alpha-1} b^{\alpha-1} E_{\text{work}}^O(j)$ . Otherwise, if  $b_j > \Gamma_j b$ , we have  $v_j < \Gamma_j^{\alpha-1} \frac{b^{\alpha-1}}{(b-1)^\alpha} E_{\text{work}}^A(W)$ . The remaining argumentation is the same as in the proof of Theorem 2.  $\square$

For  $b = \alpha + 1$  and the interesting case  $\Gamma > 1$  we get a competitive factor of  $\alpha^\alpha(1 + 2\Gamma^{\alpha-1}) + \delta_{\max} \frac{s_{\text{cr}}}{P_{\alpha,\beta}(s_{\text{cr}})}$ . For job values of at most  $\gamma$  it is  $\alpha^\alpha(1 + 2\Gamma^{\alpha-1})$ .

## 6 Conclusion & Outlook

We examined a new model that combines modern energy conservation techniques with profitability. Our results show an inherent connection between the necessary and sufficient competitive ratio of rejection-oblivious algorithms and the maximum value density. A natural question is how far this connection applies to other, more involved algorithm classes. Can we find better strategies if allowed to reject jobs even after we invested some energy, or if taking former rejection decisions into account? Such more involved rejection policies have proven useful in other models [14, 16], and we conjecture that they would do so in our setting. Other interesting directions include models for multiple processors as well as general power functions. Pruhs and Stein [16] modeled job values and deadlines in a more general way, which seems especially interesting for our profit-oriented model.

## References

- [1] S. Albers. “Algorithms for Dynamic Speed Scaling”. In: *Proc. of the 28th International Symp. on Theoretical Aspects of Computer Science (STACS)*. Schloss Dagstuhl, 2011, pp. 1–11. DOI: 10.4230/LIPIcs.STACS.2011.1.
- [2] S. Albers. “Energy-Efficient Algorithms”. In: *Comm. of the ACM* 53.5 (2010), pp. 86–96. DOI: 10.1145/1735223.1735245.
- [3] S. Albers and A. Antoniadis. “Race to Idle: New Algorithms for Speed Scaling with a Sleep State”. In: *Proceedings of the 23rd Symposium on Discrete Algorithms (SODA)*. 2012.
- [4] S. Albers, A. Antoniadis, and G. Greiner. “On Multi-Processor Speed Scaling with Migration”. In: *Proceedings of the 23rd ACM Symposium on Parallelism in Algorithms and Architectures (SPAA)*. ACM, 2011, pp. 279–288. DOI: 10.1145/1989493.1989539.
- [5] N. Bansal, T. Kimbrel, and K. Pruhs. “Speed Scaling to Manage Energy and Temperature”. In: *Journal of the ACM* 54.1 (2007), pp. 1–39. DOI: 10.1145/1206035.1206038.
- [6] N. Bansal, H.-L. Chan, K. Pruhs, and D. Katz. “Improved Bounds for Speed Scaling in Devices Obeying the Cube-Root Rule”. In: *Proceedings of the 36th International Colloquium on Automata, Languages and Programming (ICALP)*. Springer, 2009, pp. 144–155.
- [7] N. Bansal, H.-L. Chan, T.-W. Lam, and L.-K. Lee. “Scheduling for Speed Bounded Processors”. In: *Proceedings of the 35th Intl. Colloq. on Automata, Languages and Programming (ICALP)*. Springer, 2008, pp. 409–420. DOI: 10.1007/978-3-540-70575-8\_34.
- [8] P. Baptiste. “Scheduling Unit Tasks to Minimize the Number of Idle Periods: A Polynomial Time Algorithm for Offline Dynamic Power Management”. In: *Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithm (SODA)*. ACM, 2006, pp. 364–367. DOI: 10.1145/1109557.1109598.
- [9] P. Baptiste, M. Chrobak, and C. Dürr. “Polynomial Time Algorithms for Minimum Energy Scheduling”. In: *Proceedings of the 15th Annual European Symposium on Algorithms (ESA)*. Springer, 2007, pp. 136–150. DOI: 10.1007/978-3-540-75520-3\_14.

- [10] L. A. Barroso and U. Hölzle. “The Case for Energy-Proportional Computing”. In: *Computer* 40.12 (2007), pp. 33–37.
- [11] S. Baruah, G. Koren, B. Mishra, A. Raghunathan, L. Rosier, and D. Shasha. “On-line Scheduling in the Presence of Overload”. In: *Proc. of the 32nd Symp. on Foundations of Computer Science (FOCS)*. 1991, pp. 100–110. DOI: 10.1109/SFCS.1991.185354.
- [12] H.-L. Chan, T.-W. Lam, and R. Li. “Tradeoff between Energy and Throughput for Online Deadline Scheduling”. In: *Proc. of the 8th Intl. Workshop on Approximation and Online Algorithms (WAOA)*. Springer, 2010, pp. 59–70. DOI: 10.1007/978-3-642-18318-8\_6.
- [13] H.-L. Chan, W.-T. Chan, T.-W. Lam, L.-K. Lee, K.-S. Mak, and P. W. H. Wong. “Energy Efficient Online Deadline Scheduling”. In: *Proceedings of the 18th Symposium on Discrete Algorithms (SODA)*. SIAM, 2007, pp. 795–804.
- [14] X. Han, T.-W. Lam, L.-K. Lee, I. K. K. To, and P. W. H. Wong. “Deadline Scheduling and Power Management for Speed Bounded Processors”. In: *Theoretical Computer Science* 411.40–42 (2010), pp. 3587–3600. DOI: 10.1016/j.tcs.2010.05.035.
- [15] S. Irani, S. Shukla, and R. Gupta. “Algorithms for Power Savings”. In: *ACM Transactions on Algorithms* 3.4 (2007). DOI: 10.1145/1290672.1290678.
- [16] K. Pruhs and C. Stein. “How to Schedule When You Have to Buy Your Energy”. In: *Proc. of the 13th/14th Workshop on Approximation Algorithms for Comb. Optimization Problems/Randomization and Computation (APPROX/RANDOM)*. Springer, 2010, pp. 352–365. DOI: 10.1007/978-3-642-15369-3\_27.
- [17] F. F. Yao, A. J. Demers, and S. Shenker. “A Scheduling Model for Reduced CPU Energy”. In: *Proceedings of the 36th Annual Symposium on Foundations of Computer Science (FOCS)*. 1995, pp. 374–382.